Statement of Work (SOW)

# Project Title: Smart Perishable Demand Forecasting in Grocery Retail

## 1. Project Overview

This project aims to address the inventory challenges faced by grocery retailers due to the highly variable demand of perishable goods such as milk, fruits, vegetables, and meat. These challenges are influenced by external factors like weather, calendar effects, and store-specific dynamics. The goal is to reduce waste and avoid stockouts by building a robust demand forecasting system.

## 2. Objectives

* Develop a predictive model to forecast daily demand at the SKU-store level.
* Integrate weather, calendar, and product type variables to improve accuracy.
* Enable data-driven replenishment decisions for grocery retailers.
* Build visual dashboards for insight-driven decision-making.

## 3. Deliverables

* A cleaned and pre-processed dataset (Data\_for\_2024.csv) with key features like temperature, rainfall, date, store, and product type.
* A Python-based forecasting pipeline using LightGBM to predict daily unit sales.
* An interactive user input script to simulate predictions based on store ID, product type, temperature, and rainfall.
* A Tableau dashboard to visualize trends, seasonality, and demand forecasts.
* Full project documentation outlining methodology and results.

## 4. Methodology (Python Implementation)

* Imported relevant libraries (Pandas, Seaborn, LightGBM, etc.).
* Loaded and explored the dataset (head, info, describe, missing values).
* Converted date columns to datetime format.
* Checked and treated outliers in weather variables (temperature, rainfall) using boxplots.
* Generated distribution plots for numeric features.
* Filtered out extreme values to ensure model reliability.
* Trained a LightGBM regression model and evaluated using MAE and RMSE.
* Developed a prediction interface for store-level use cases.

## 5. Business Value

* Minimize spoilage from overstocking.
* Reduce customer dissatisfaction due to stockouts.
* Optimize shelf space and inventory costs.
* Improve overall store performance through data-driven decisions.